Covid 19

Abanoub George, Marc Ashraf, Alaa Yehia, Salma Mohamed, Omar Badry

June 2020

Department of Computer Science, Misr International University (MIU)

Email: abanoub1703340@miuegypt.edu.eg

Email: mark1712199@miuegypt.edu.eg

Email: alaa1709041@miuegypt.edu.eg

Email: salma1711570@miuegypt.edu.eg

Email: omar1707863@miuegypt.edu.eg

GitHub: https://github.com/BonyGeorge/Artificial-Intelligence-Project-

Contents

List of Figures

1 Abstract

COVID-19 was first established in Wuhan city in China, then it spreaded worldwide to alot of countries. The number of infected countries differ from one to another according to the medications and the number of population. To help us know the number of infected people of COVID-19 in each country, we used two different datasets and four different algorithms which are KNN, SVM, K means, Decision tree. .The first dataset used was about countries and the number predicting infected patients and the real number of infected patients of the current time,this dataset was used in KNN and Decision tree algorithms. KNN algorithm was used by Python. The second dataset used was about 210 countries with their continent, population, total cases, total deaths, total recovered, new death and new recovered. This dataset was used in SVM and K-means algorithms. SVM, K-means, Decision tree algorithms was used by Weka.

1.1 Keywords

COVID-19, KNN, K-means, SVM, Decision tree, Python, Weka.

2 Introduction

COVID-19 virus was first established in Wuhan city in China in December 2019 and it quickly spread worldwide. In March 2020, Most of the infected countries closed their airlines and factories and the education had stopped due to the huge spreadness of Corona virus. As of Feb 24, 2020, there was about 80,000 approved cases have been announced in more than 28 countries.[3][8][4][2] To this moment, No drugs or vaccine or supplement have been detected to get rid of the virus from the body.

The used algorithms in this project are KNN, SVM, K means, Decision tree.

1. KNN: It resembles K nearest neighbors in which it stores all obtainable cases then it distinguishes new cases based on similarity measures.It is responsible of recognition and regression.

2. SVM: SVM algorithm is used in classification and regression problems. In SVM algorithm, each data item is set as a point in n-dimensional space, with the value of each feature being the value of a particular coordinate.

3. K means: K means algorithm is considered as a unsupervised algorithm. It is also considered a method of vector quantization.K means work as follow, First it identifies the center of cluster,Then it assigns the nearest cluster to data point.Then setting the position of each cluster.

4. Decision tree: Decision tree algorithm is used for solving classification and regression problems. The purpose of using this algorithm is to create a trained dataset which will be used to know the value of the target.

Our goal in this document is very simple, which is to compare between the number of infected people in each country, and then study how every country held this situation. Although none of us have a medical background, We tried our best to study the situation of COVID-19 in each country and its numbers.

3 Literature survey

Alot of researchers made researches on this issue especially the issue is spread worldwide and it is still brand new. The first research [7] was made by Max Roser, Hannah Ritchie and Esteban Ortiz-Ospina.This research discusses the statistics of the spread of COVID-19 which takes its information from WHO(world health organization).The second research [5] was made by Mary A Lake in March 2020 and she talked about the current knowledge of COVID-19. The third research [1] was made by (Mohamad Chahrour, Sahar Assi, Michael Bejjani, Ali A Nasrallah, Hamza Salhab, Mohamad Fares, and Hussein H Khachfe) which talked about COVID-19 since its discovery since December 2019. The last research [6] talked about the raise of number of corona virus worldwide and its symptoms.

4 Proposed COVID-19 system

We used two different datasets as we used many different algorithms.

4.1 First Dataset

First the dataset we used is a dataset of countries with its infected patients in each country .This dataset was made recently. This dataset was used for KNN algorithm. These datasets is based on when entering a test data, it compares the real data (test data) with the predicted ones which was already in the dataset.

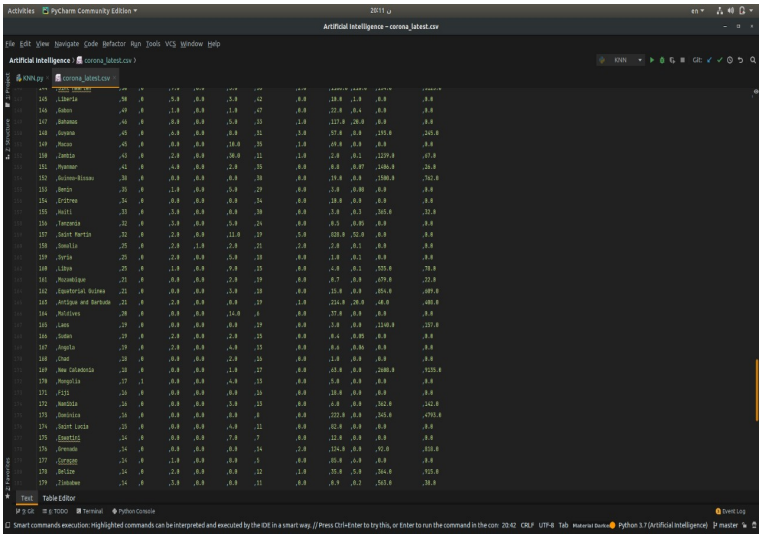

Figure 1: First Dataset Part 1

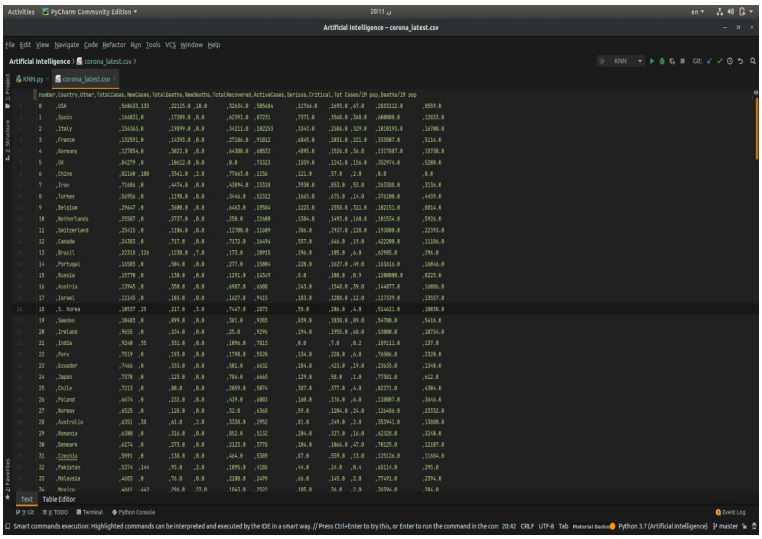

Figure 2: First Dataset Part 2

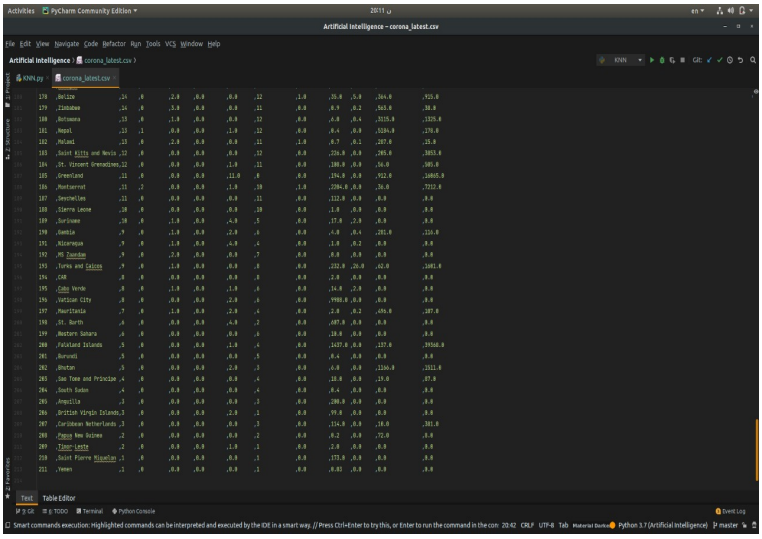

Figure 3: First Dataset Part 3

4.1.1 KNN algorithm using first dataset

KNN Algorithm : it stands for K Nearest Neighbors where the K is a constant number that we define to to search for the nearest data in a distance that we specify by using different methods such as Euclidean and Manhattan methods to get the distance between two points. We used here the KNN to test the data

which is a csv file and then we gave a label to each column which is a number in our data set then we tested it and got the predicted data of how many people in each country is sick with COVID-19 and the actual data which is in the data set and also we calculated the distances between them . Here the K we used is 9 to get the nearest 9 neighbors.

This was the code used using KNN algorithm .

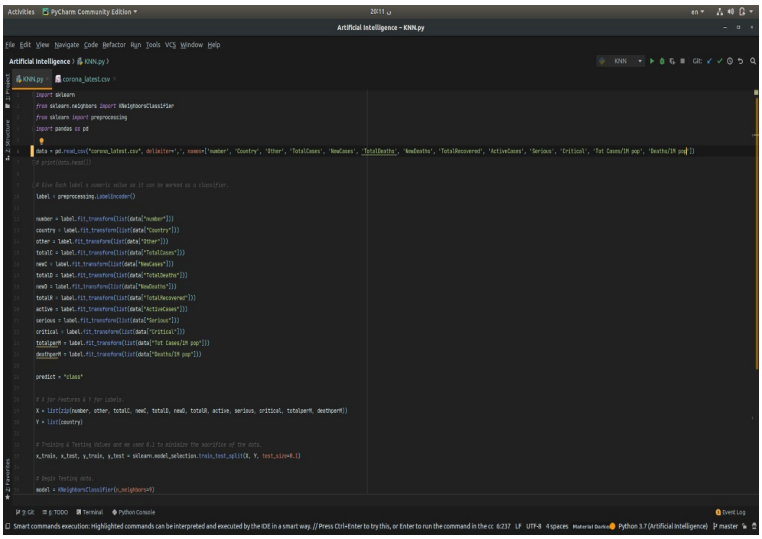

Figure 4: KNN algorithm code Part 1

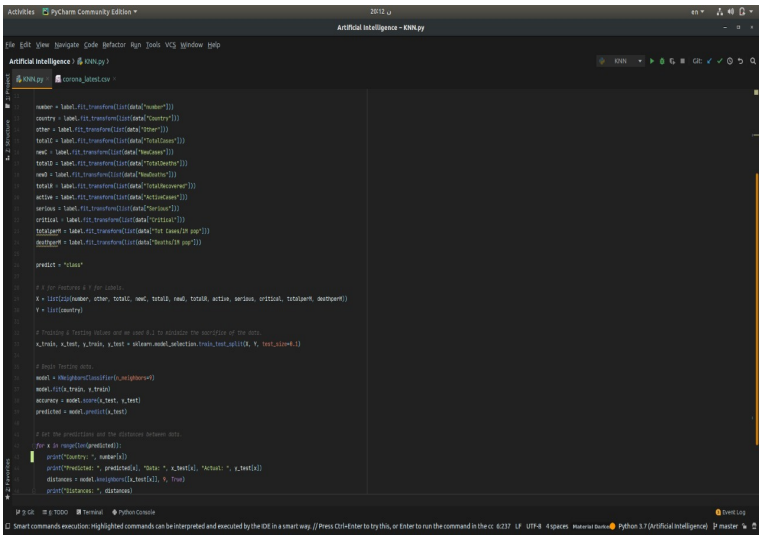

Figure 5: KNN algorithm code Part 2

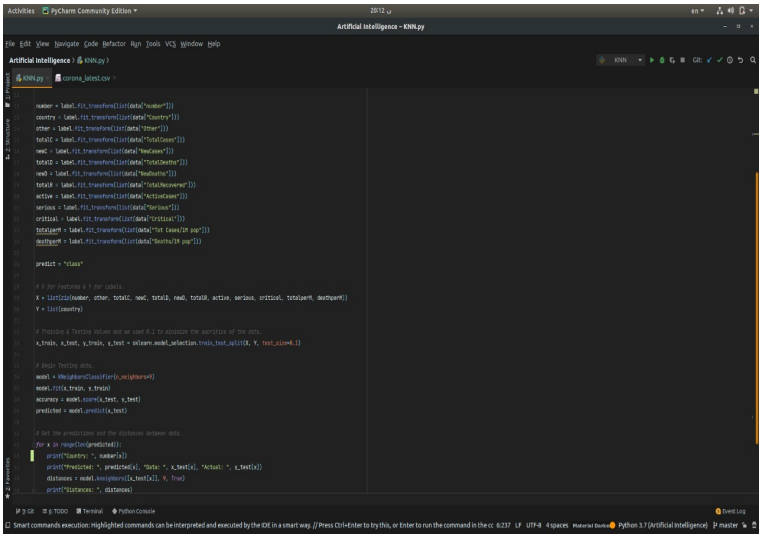

Figure 6: KNN algorithm code Part 3

4.1.2 Decision tree using First Dataset

Decision Tree : this algorithm depends on dividing the dataset into sub classes and keep doing this method till it reach the classifier it is one of the supervised algorithms. Here we used it to divide each country with it's own data from deaths, cases and critical cases.

4.2 Second Dataset

The second dataset used consists of 210 country and their continents.Each country has its own data which are population, total cases, total death, new death, total recovered, new recovered, acive cases, serious/critital cases.

Country/Region	Continent	Population	TotalCases	NewCases	TotalDeaths	NewDeaths	TotalRecovered	NewRecovered	ActiveCases	Serious,Critical	Tot Cases/1M pop	Deaths/1M pop	TotalTests	Tests/1M pop
USA	North Am	3.31E+08	2007449		112469		761708		1133272	16923	6067	340	21291677	64349
Brazil	South Am	2.12E+08	691962		37312	813	302084		352566	8318	3257	176	999836	4706
Russia	Europe	1.46E+08	467673		5859		226731		235083	2300	3205	40	12721549	87175
Spain	Europe	46753690	288630		27136					617	6173	580	4063843	86920
UK	Europe	67863688	286194		40542					604	4217	597	5581073	82239
India	Asia	1.38E+09	258090	604	7207		124095	247	126788	8944	187	5	4774434	3462
Italy	Europe	60467085	234998		33899		165837		35262	287	3886	561	4236535	70063
Peru	South Am	32941004	196515		5465		86219		104831	1062	5966	166	1191956	36185
Germany	Europe	83767456	185869		8776		169600	500	7493	568	2219	105	4348880	51916
Iran	Asia	83921387	171789		8281		134349		29159	2596	2047	99	1084857	12927
Turkey	Asia	84279639	170132		4692		137969		27471	613	2019	56	2338593	27748
France	Europe	65264696	153977		29155		70842		53980	1053	2359	447	1384633	21216
Chile	South Am	19105644	134150		2190		95631		36329	1558	7021	115	708773	37098
Mexico	North Am	1.29E+08	117103	3484	13699	188	84419	2875	18985	378	909	106	336395	2611
Pakistan	Asia	2.21E+08	103671	4728	2067	65	34355	890	67249	111	470	9	705833	3200
Saudi Arabia	Asia	34776977	101914		712		72817		28385	1564	2931	20	958237	27554
Canada	North Am	37720832	95699		7800		54233		33666	1816	2537	207	1896822	50286
Qatar	Asia	2807805	68790		54		44338		24398	245	24500	19	255533	91008
Bangladesh	Asia	1.65E+08	65769		888		13903		50978	1	400	5	397987	2418
Belgium	Europe	11586489	59226		9595		16291		33340	111	5112	828	935670	80755
Belarus	Europe	9449509	48630		269		23647		24714	92	5146	28	622313	65857
South Africa	Africa	59258991	48285		998		24364		22923	208	815	17	920064	15526
Netherlands	Europe	17132558	47574		6013					97	2777	351	388113	22654
Sweden	Europe	10095291	44730		4659					198	4431	462	275500	27290
Ecuador	South Am	17624879	43120		3621		21020		18479	222	2447	205	127414	7229
Colombia	South Am	50847406	39236		1259		15322		22655	335	772	25	410719	8077
UAE	Asia	9882493	38808		276		21806		16726	1	3927	28	2500000	252973
Singapore	Asia	5847399	37910		25		24886		12999	3	6483	4	408495	69859
Portugal	Europe	10198446	34693		1479		20995		12219	58	3402	145	873998	85699
Egypt	Africa	1.02E+08	34079		1237		8961		23881	41	333	12	135000	1321
Kuwait	Asia	4266300	31848		264		20205		11379	196	7465	62	315285	73901
Indonesia	Asia	2.73E+08	31186		1851		10498		18837		114	7	405992	1485
Switzerland	Europe	8650592	30965		1921		28700		344	28	3580	222	423468	48952
Ukraine	Europe	43748704	26999		788		12054		14157	304	617	18	424046	9693
Poland	Europe	37849071	26561		1157		12855		12549	160	702	31	1056396	27911
Ireland	Europe	4934162	25201		1679		22698		824	36	5107	340	348416	70613
Argentina	South Am	45168963	22794		664		6909		15221	274	505	15	193923	4293
Philippines	Asia	1.09E+08	21895		1003		4530		16362	82	200	9	430081	3928
Afghanistan	Asia	38865373	20917	575	369	12	2171	296	18377	19	538	9	48305	1243

Figure 7: Second Dataset Part 1

Romania	Europe	19244947	20479		1333		14638		4508	144	1064	69	503200	26147
Dominican Republic	North Am	10840829	19600		538		12007		7055	110	1808	50	94511	8718
Israel	Asia	9197590	17915	52	298		15102	11	2515	27	1948	32	662916	72075
Japan	Asia	1.26E+08	17141		916		15139		1086	98	136	7	314483	2486
Austria	Europe	9003169	16902		672		15793		437	17	1877	75	489597	54381
Oman	Asia	5097088	16882		75		3451		13356	75	3312	15	108679	21322
Panama	North Am	4310139	16425		393		10218		5814	84	3811	91	72697	16867
Bahrain	Asia	1696917	14763		26		9468		5269	13	8700	15	367056	216308
Bolivia	South Am	11662351	13643	285	465	11	2086	184	11092	3	1170	40	36530	3132
Armenia	Asia	2962906	13130		200		4014		8916	10	4431	68	69171	23346
Kazakhstan	Asia	18761857	12859	165	56	3	7376		5427	62	685	3	965089	51439
Nigeria	Africa	2.06E+08	12486		354		3959		8173	7	61	2	76802	373
Iraq	Asia	40157701	12366		346		5186		6834	69	308	9	303053	7547
Denmark	Europe	5790943	11948		589		10755		604	15	2063	102	706442	121991
Serbia	Europe	8739407	11823		249		11348		226	15	1353	28	272502	31181
S. Korea	Asia	51266523	11814	38	273		10563	11	978	15	230	5	1018214	19861
Algeria	Africa	43796100	10154		707		6717		2730	24	232	16		
Moldova	Europe	4034514	9700		341		5638		3721	335	2404	85	63328	15697
Ghana	Africa	31026938	9638		44		3636		5958	3	311	1	233734	7533
Czechia	Europe	10707772	9628		327		6891		2410	12	899	31	471521	44035
Norway	Europe	5418531	8547		238		8138		171	5	1577	44	257303	47486
Malaysia	Asia	32338393	8322		117		6674		1531	5	257	4	610038	18864
Morocco	Africa	36881658	8224		208		7364		652	15	223	6	305953	8296
Cameroon	Africa	26497531	7908		212		4735		2961	28	298	8		
Azerbaijan	Asia	10133277	7553		88		4149		3316	66	745	9	328385	32407
Australia	Australia/	25480378	7265	5	102		6706	3	457	3	285	4	1632116	64054
Guatemala	North Am	17892493	7055	263	252	22	1261	128	5542	5	394	14	31427	1756
Finland	Europe	5540198	6981		323		5800		858	6	1260	58	201000	36280
Honduras	North Am	9893859	6327	172	258	8	712	15	5357	13	639	26	21540	2177
Sudan	Africa	43775253	6081		359		2014		3708		139	8	401	9
Tajikistan	Asia	9522265	4529		48		2673		1808		476	5		
Uzbekistan	Asia	33436477	4352	21	17		3354		981	7	130	0.5	748555	22387
Senegal	Africa	16711294	4328		49		2588		1691	15	259	3	49180	2943
Djibouti	Africa	987030	4207		28		1877		2302		4262	28	35165	35627
Guinea	Africa	13106296	4117		23		2877		1217	24	314	2	14407	1099
Luxembourg	Europe	625281	4039		110		3899		30	1	6459	176	88051	140818
DRC	Africa	89353804	4016		85		537		3394		45	1		
Hungary	Europe	9661789	4008		546		2279		1183	21	415	57	210202	21756
Ivory Coast	Africa	26330544	3739		36		1818		1885		142	1	31549	1198
Nepal	Asia	29100465	3448		13		467		2968		118	0.4	241254	8290

Figure 8: Second Dataset Part 2

Romania	Europe	19244947	20479		1333		14638		4508	144	1064	69	503200	26147
Dominican Republic	North Am	10840829	19600		538		12007		7055	110	1808	50	94511	8718
Israel	Asia	9197590	17915	52	298		15102	11	2515	27	1948	32	662916	72075
Japan	Asia	1.26E+08	17141		916		15139		1086	98	136	7	314483	2486
Austria	Europe	9003169	16902		672		15793		437	17	1877	75	489597	54381
Oman	Asia	5097088	16882		75		3451		13356	75	3312	15	108679	21322
Panama	North Am	4310139	16425		393		10218		5814	84	3811	91	72697	16867
Bahrain	Asia	1696917	14763		26		9468		5269	13	8700	15	367056	216308
Bolivia	South Am	11662351	13643	285	465	11	2086	184	11092	3	1170	40	36530	3132
Armenia	Asia	2962906	13130		200		4014		8916	10	4431	68	69171	23346
Kazakhstan	Asia	18761857	12859	165	56	3	7376		5427	62	685	3	965089	51439
Nigeria	Africa	2.06E+08	12486		354		3959		8173	7	61	2	76802	373
Iraq	Asia	40157701	12366		346		5186		6834	69	308	9	303053	7547
Denmark	Europe	5790943	11948		589		10755		604	15	2063	102	706442	121991
Serbia	Europe	8739407	11823		249		11348		226	15	1353	28	272502	31181
S. Korea	Asia	51266523	11814	38	273		10563	11	978	15	230	5	1018214	19861
Algeria	Africa	43796100	10154		707		6717		2730	24	232	16		
Moldova	Europe	4034514	9700		341		5638		3721	335	2404	85	63328	15697
Ghana	Africa	31026938	9638		44		3636		5958	3	311	1	233734	7533
Czechia	Europe	10707772	9628		327		6891		2410	12	899	31	471521	44035
Norway	Europe	5418531	8547		238		8138		171	5	1577	44	257303	47486
Malaysia	Asia	32338393	8322		117		6674		1531	5	257	4	610038	18864
Morocco	Africa	36881658	8224		208		7364		652	15	223	6	305953	8296
Cameroon	Africa	26497531	7908		212		4735		2961	28	298	8		
Azerbaijan	Asia	10133277	7553		88		4149		3316	66	745	9	328385	32407
Australia	Australia/	25480378	7265	5	102		6706	3	457	3	285	4	1632116	64054
Guatemala	North Am	17892493	7055	263	252	22	1261	128	5542	5	394	14	31427	1756
Finland	Europe	5540198	6981		323		5800		858	6	1260	58	201000	36280
Honduras	North Am	9893859	6327	172	258	8	712	15	5357	13	639	26	21540	2177
Sudan	Africa	43775253	6081		359		2014		3708		139	8	401	9
Tajikistan	Asia	9522265	4529		48		2673		1808		476	5		
Uzbekistan	Asia	33436477	4352	21	17		3354		981	7	130	0.5	748555	22387
Senegal	Africa	16711294	4328		49		2588		1691	15	259	3	49180	2943
Djibouti	Africa	987030	4207		28		1877		2302		4262	28	35165	35627
Guinea	Africa	13106296	4117		23		2877		1217	24	314	2	14407	1099
Luxembourg	Europe	625281	4039		110		3899		30	1	6459	176	88051	140818
DRC	Africa	89353804	4016		85		537		3394		45	1		
Hungary	Europe	9661789	4008		546		2279		1183	21	415	57	210202	21756
Ivory Coast	Africa	26330544	3739		36		1818		1885		142	1	31549	1198
Nepal	Asia	29100465	3448		13		467		2968		118	0.4	241254	8290

Figure 9: Second Dataset Part 3

4.2.1 K-means algorithm using second dataset

K-means : This algorithm is used for unsupervised learning in which it clusters the data by grouping the clusters comparable data. The K here is constant we choose it. Here we used it to to group all the data for each country .

4.2.2 Support vector machines algorithm using second dataset

SVM: stands for " Support vector machines " it separates 2 classes by a straight line (can be viewed the visualizer). a device called "kernel trick" used when the classes cannot be separated to make boundaries of different shaped.

5 Result and Discussion

5.1 KNN algorithm

The results of KNN algorithm appears as follows, as all countries are retrieved with their numbers only but not their names. Each country retrieve the data of its predicted number of infected patients and the real number of infected patients.

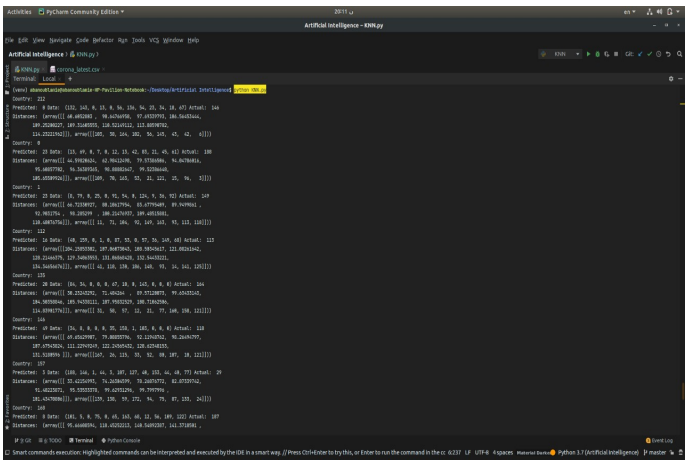

Figure 10: KNN Results Part 1

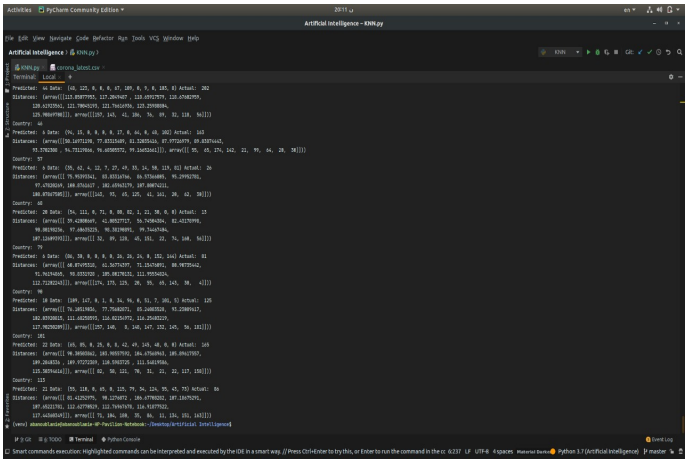

Figure 11: KNN Results Part 2

5.2 Decision tree algorithm

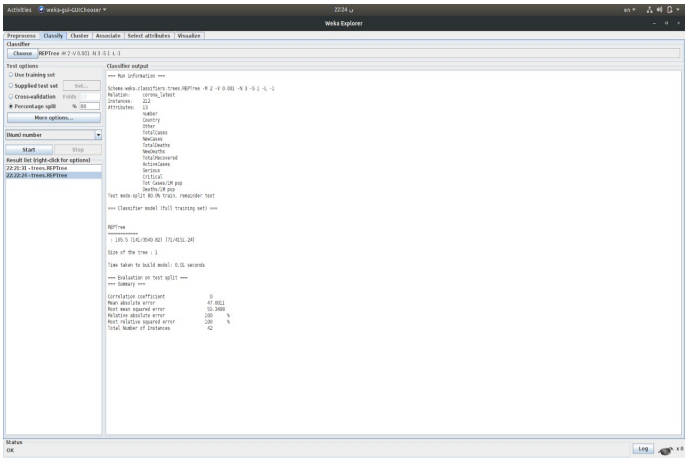

Figure 12: Decision tree algorithm results with split 80%

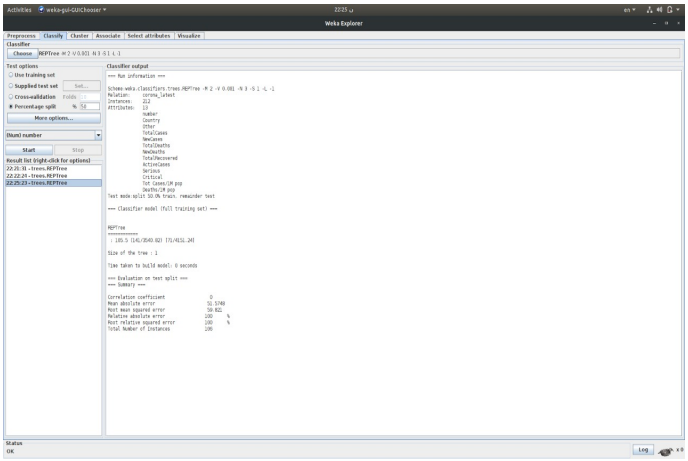

Figure 13: Decision tree algorithm results with split 50%

5.3 K-means Algorithm

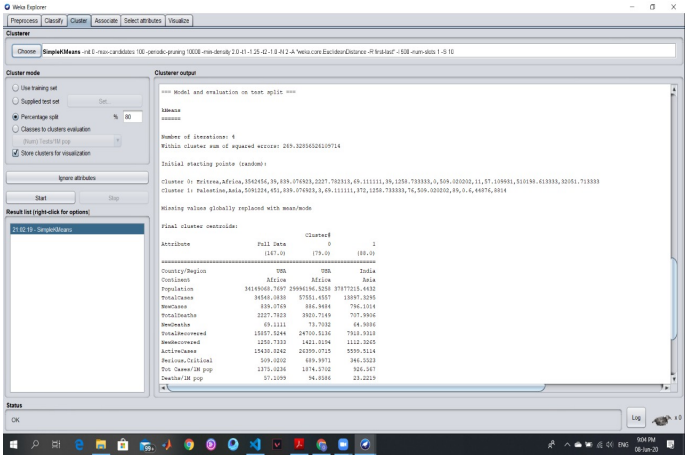

Figure 14: K-means results Part 1

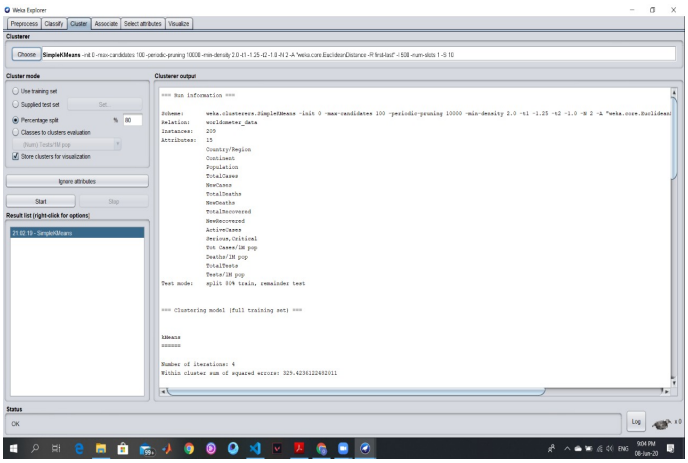

Figure 15: K-means results Part 2

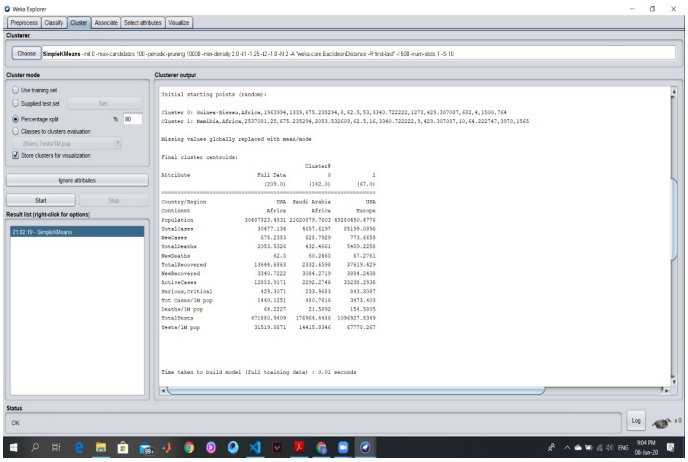

Figure 16: K-means results Part 3

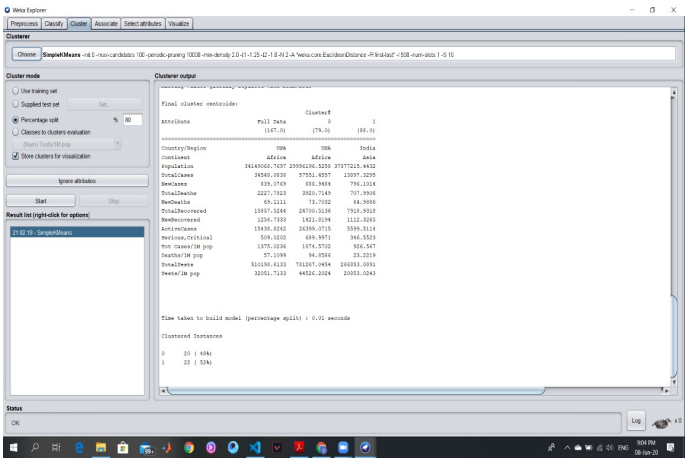

Figure 17: K-means results Part 4

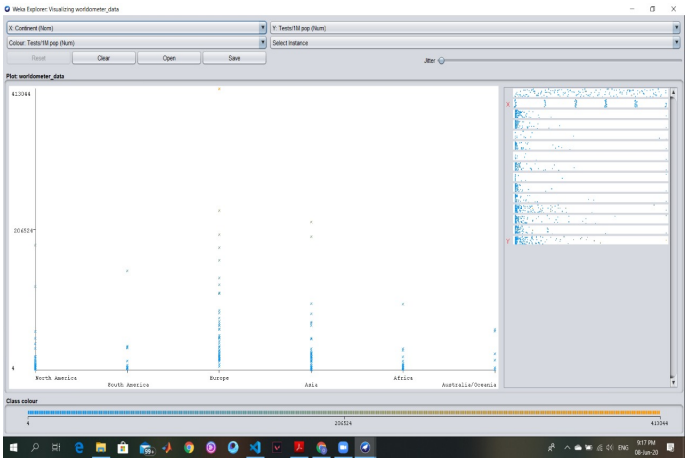

Figure 18: K-means results visualize

5.4 SVM algorithm

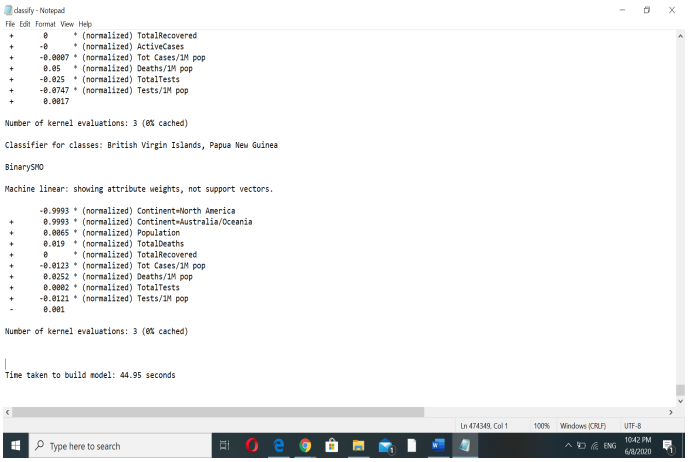

Figure 19: SVM results : Time taken to build model

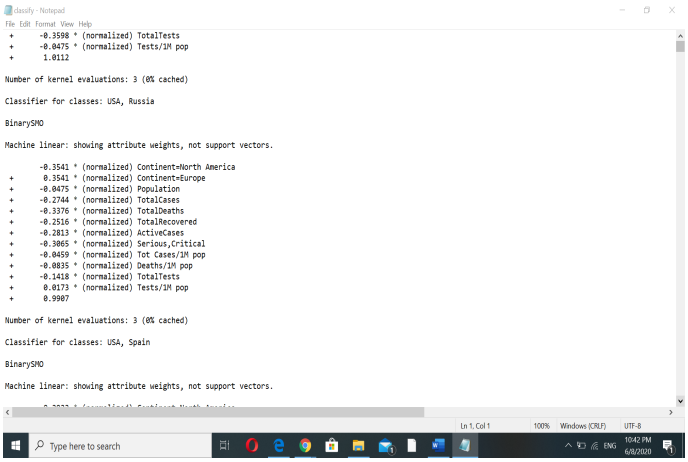

Figure 20: SVM results : Comparing between USA and Russia

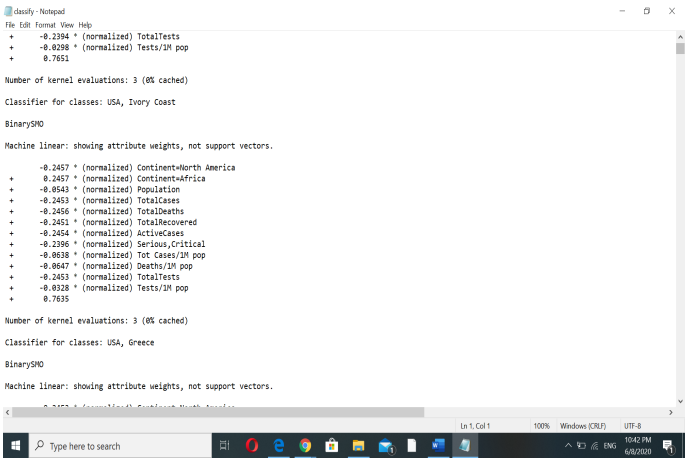

Figure 21: SVM results : Comparing between USA and Ivory Coast

6 Conclusion

This document was made to discuss the difference between the number of infected patients between all countries. We wanted to facilitate this process by using some algorithms which are KNN, SVM, K means and Decision tree. KNN algorithm was used by Python, while SVM, K-means and Decision tree algorithms was used with Weka.

References

[1] Mohamad Chahrour et al. "A bibliometric analysis of Covid-19 research activity: A call for increased output". In: *Cureus* 12.3 (2020).

[2] Yawei Chang and Houquan Liu. "Semi-supervised classification algorithm based on the KNN". In: (2011), pp. 9–12.

[3] Chenlin GU et al. "Mathematical recommendations to fight against COVID-19". In: *Available at SSRN 3551006* (2020).

[4] Nirmal Kandel et al. "Health security capacities in the context of COVID-19 outbreak: an analysis of International Health Regulations annual report data from 182 countries". In: *The Lancet* (2020).

[5] Mary A Lake. "What we know so far: COVID-19 current clinical knowledge and research". In: *Clinical Medicine* 20.2 (2020), p. 124.

[6] Canelle Poirier et al. "The Role of Environmental Factors on Transmission Rates of the COVID-19 Outbreak: An Initial Assessment in Two Spatial Scales." In: *Available at SSRN 3552677* (2020).

[7] Max Roser et al. "Coronavirus disease (COVID-19)–Statistics and research". In: *Our World in data* (2020).

[8] Fei Zhou et al. "Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study". In: *The lancet* (2020).

7 Other References

1. https://www.saedsayad.com/k_nearest_neighbors.htm#: :text=K%20nearest%20neighbors%20is%20a,as%20a%20non%2Dparametric%20technique
2. https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/
3. https://www.edureka.co/blog/k-means-clustering/
4. http://www.onmyphd.com/?p=k-means.clustering
5. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200225-sitrep-36-covid-19.pdf?sfvrsn=2791b4e0_2
6. https://www.kaggle.com/balaaje/coronavirus-covid19-dataset
7. https://www.kaggle.com/imdevskp/corona-virus-report